Crystal structure of the *Arabidopsis thaliana* L1L/NF-YC3 histone-fold dimer reveals specificities of the LEC1 family of NF-Y subunits in plants.

Nerina Gnesutta[1], Dana Saad[1], Antonio Chaves-Sanjuan[1], Roberto Mantovani[1], Marco Nardini[1,*]

[1]Dipartimento di Bioscienze, Università degli Studi di Milano, Via Celoria 26, 20133, Milano, Italy.

***Correspondence: Marco Nardini** (marco.nardini@unimi.it)

Running Title: X-ray structure of *Arabidopsis* L1L/NF-YC3 dimer

Short Summary

The structural analysis of the L1L/NF-YC3 histone dimer from *A. thaliana* shows that it can trimerize and bind to the CCAAT box. Specific sequence and structural features at the protein/DNA interface, in particular the presence of a Asp-His pair, help explain the molecular mechanisms of LEC1/L1L activity as a *bona fide* mammalian-like NF-YB.

Dear Editor,

the NF-Y transcription factor is a heterotrimer formed by evolutionarily conserved subunits: NF-YA, NF-YB, and NF-YC. NF-YB and NF-YC harbor a Histone Fold Domain (HFD), structurally similar to that of nucleosome core histones, and form a tight dimer (Romier et al., 2003). NF-YA binds to the NF-YB/NF-YC dimer and provides exquisite sequence-specificity for recognition and binding of the CCAAT box (Nardini et al., 2013; Huber et al., 2012), an important DNA regulatory element of all eukaryotes (Dolfini et al., 2009). In plants, each NF-Y subunit is expanded to 8-14 genes (Laloum et al., 2012). Within the NF-YB genes, *LEAFY COTYLEDON1* (*LEC1)* and *LEC1-LIKE* (*L1L*) form a conserved subfamily (Braybrook and Harada, 2008), originally identified in genetic experiments in *Arabidopsis thaliana* (At-*LEC1*, corresponding to At-*NF-YB9*, and At-*L1L*, corresponding to At-*NF-YB6*, respectively) for their key roles in embryo maturation (Lee et al., 2003, Kwong et al., 2003). Considerations based on available structures and sequence alignments of At-NF-YBs raise questions as to LEC1/L1L capacity to bind DNA. In addition, activation of seed specific genes by L1L, such as *CRC*, *SUS2* and the lipidogenic *FAD3,* involves indirect promoter tethering by interactions with bZIP67, binding to Abscisic Acid Response Elements (ABREs) (Yamamoto et al., 2009; Mendes et al., 2013).

To understand the specificities of the plant LEC1-like NF-YBs it is mandatory to unveil their structural features. To this aim, we crystallized and solved the structure at 2.3 Å resolution of the core domains of At-L1L and At-NF-YC3 subunits (See Supplemental Materials and Methods, Table S1, and Figures S1A and S2). As expected, both subunits adopt a HFD structure (helix α1-loop L1-helix α2-loop L2-

helix α3) and interact in a head-to-tail fashion, forming a classical histone-like pair (Figure 1A). Both At-L1L and At-NF-YC3 show the typical structural features specific for NF-YB and NF-YC subunits, as highlighted by sequence/structure comparison with mammalian NF-YB/NF-YC heterodimer (sequence identity of 63.4% for NF-YB and 73.8% for NF-YC; r.m.s.d. of 0.95 Å calculated over 170 Cα pairs) (Figure 1A) including, among others, the presence of an intra-chain Arg-Asp bidentate pair in both subunits (Arg114-Asp121 in At-L1L, and Arg121-Asp128 in At-NF-YC3), and the presence of a Trp residue, absolutely conserved in NF-YCs (Trp113 in At-NF-YC3), nestled in a hydrophobic pocket at the interface between the L1/L2 loops pair of NF-YB and NF-YC (see Figures S2 and S3). The only divergent regions that distinguish the protein backbone of At-L1L/NF-YC3 from that of mammalian NF-YB/NF-YC heterodimer are localized at the N- and C-terminus (Figure 1A).

We then analysed the surface of the At-L1L/NF-YC3 dimer in search for the presence of positively charged residues in the in α1, L1 and L2 regions, predicted to be involved in histone-like DNA-contacts (see Figure S2). Figure 1B (top view) shows that, bar minor lateral differences essentially related to At-L1L residues Asp84 in helix α2 and His79 in loop L1 (discussed below), the overall surface of the At-L1L/NF-YC3 dimer is very similar to mammalian NF-YB/NF-YC, thus perfectly suited for binding and bending the DNA (Nardini et al., 2013). We noticed also the presence of a wide, acidic surface involving the At-L1L helix α2 and the At-NF-YC3 helix αC, which matches perfectly the location of the NF-YA binding groove in NF-YB/NF-YC (Figure 1B, side view). Thus, *Arabidopsis* NF-YAs are expected to bind the HFD heterodimer as mammalian NF-YA does. Indeed, almost the totality of the

mammalian HFD residues interacting with NF-YA and with DNA are conserved in At-L1L/NF-YC3 (see Figure S2).

We wished to ascertain whether trimers composed of plant NF-Y subunits indeed bind DNA. For this purpose, we performed EMSA experiments, using the At-L1L/NF-YC3 HFD dimer and the At-NF-YA6 subunit, widely expressed in seeds. These experiments extend the information previously published (Calvenzani et al., 2012), where At-L1L trimerization and DNA binding was tested in EMSA by using recombinant mouse NF-YA and NF-YC.

In the absence of genetic information on a CCAAT-containing promoter expressed in seeds, hence potentially targeted by our At-NF-Y trimer, we used a fluorescently labelled high affinity CCAAT box probe derived from the human HSP70 promoter (See Supplemental Materials and Methods). EMSA analysis with increasing doses of the HFD heterodimer or At-NF-YA6, in the presence of the respective counterpart, shows robust DNA binding only in the presence of At-L1L/NF-YC3/NF-YA6 trimers and not of the separated subunits (Figure 1C). Furthermore, the bands corresponding to the At-NF-Y/DNA complex are specific for CCAAT, since it is competed away by an unlabelled CCAAT oligo, but not by one with an unrelated sequence (Figure 1D). The Asn67→Asp mutation in helix α1 of At-L1L, a key contact site between NF-YB and the DNA phosphate backbone in the mammalian NF-Y/DNA complex (Sinha et al., 1996; Zenzoumi et al., 1999; Nardini et al., 2013), abrogates the ability of At-L1L/NF-YC3 to bind DNA in EMSA experiments, without affecting heterodimerization. A similar result is obtained with the Glu96→Arg mutation in helix α2 of At-L1L, a site demonstrated to be diagnostic for NF-YA association and trimerization (Sinha et al., 1996) (Figures S1C and S4A). Altogether, the data indicate

that At-L1L and At-NF-YC3 are *bona-fide* NF-Y subunits able to form a canonical HFD heterodimer, with a trimerization and a DNA binding mode conforming to the mammalian NF-Y.

Nevertheless, significant differences also emerge. Among the HFD residues involved in DNA-interactions Asp84 of At-L1L (Asp55 in LEC1) stands out, since it is typically Lys/Arg in other plant, yeast and mammalian NF-YBs (Figure S2A). In mammalian NF-YB, the corresponding Lys78, located in helix α2, contacts the DNA phosphate backbone, 2 bps upstream of CCAAT (Nardini et al., 2013). This residue was a major focus of our attention, since genetic experiments pinpointed Asp55 as crucial for LEC1 function *in vivo*: an Asp→Lys mutation led to a dramatic loss of LEC1 activity, while a Lys→Asp substitution was sufficient to confer partial LEC1 features to the more "canonical" At-NF-YB3 (Lee et al., 2003). EMSA experiments on At-L1L wt and Asp84→Lys mutant show that both proteins bind the DNA with similar affinities (Figures S1C, S4A). The current dimer structure, superimposed on the quaternary NF-Y/CCAAT complex (Figure 1E), suggests that the electrostatic repulsion between the negatively charged Asp84 and the DNA phosphate backbone would favor a slightly shifted DNA trajectory, stabilized on the opposite side of the double helix by His79, which could compensate the missing DNA-contact resulting from the Lys→Asp substitution at position 84 in At-L1L (or 55 in LEC1) (Figure 1E). Notably, this His residue is absolutely conserved in plants, and diagnostic for LEC1 family members (see Figure S2A). Note that the corresponding mammalian Thr73 is not in contact with DNA (Figure 1E), and essentially all "canonical" NF-YBs in plants have an Asn residue at this position (see Figure S2A). In summary, the At-L1L model posits that Asp84 would "shift" DNA toward His79, making contacts with the

complementary strand, also upstream of CCAAT. This structural variability in the protein-DNA interactions at the At-L1L Asp84 site (Lys in standard NF-YB subunits), supports the idea that this protein region in different NF-YB subunits might provide selectivity for different composition and flexibility of GC *vs* AT sequences flanking the CCAAT nucleotides, thus accounting for the functional differences in gene activation that distinguish members of the L1L/LEC1-type subfamily from other NF-YBs.

An additional difference is observed at the N-terminal parts of At-L1L and At-NF-YC3, both more extended relative to their mammalian counterparts (Figure 1A). In the crystal unit cell, the At-L1L N-terminus inserts between the core domain and the N-terminal region of an adjacent At-NF-YC3 molecule (see Figure S5A). Formally, it is possible that crystal growth/packing select for the extended conformation of the N-termini of both subunits, and that the physiological conformations would be different. The following considerations, however, suggest otherwise. The N-terminus of At-NF-YC3 extends differently from the short stretch of NF-YC (Figure 1A), but superimposition with NF-Y from *A. nidulans* (termed Hap: Huber et al., 2012) reveals a similar backbone path (see Figure S5B): in HapE (corresponding to NF-YC), this stretch leads to the long αN helix that runs antiparallel to, and contacts the α2 helix. Plant NF-YCs show conservation in this area (Huber et al., 2012; see caption of Figure S2B), hinting at the presence of an αN, although not to the full extension of HapE. On the At-L1L side, superimposition of the extended N-terminus reveals that it is located in the region occupied by the A1A2-linker of NF-YA (and HapB) within the NF-Y/DNA complex (Figure 1F). For this to be physiologically relevant, one would have to assume that the At-NF-YA A1A2-linker follows a different pathway due to specific amino acid composition: indeed this is an area of relative divergence

across kingdoms, and paralogs. In mammalian NF-YA, two residues, Gly260 and Pro263 guide the structure of the A1A2-linker by providing flexibility and directionality, respectively, needed to orient the A2 helix towards the CCAAT box site (Figure 1F). These two residues are absent in plant NF-YAs, including At-NF-YA6. Furthermore, in the At-NF-YA2 homolog, the A1A2-linker has four additional residues, suggesting a specific structuring of this region (see Figure S2C). Independently on the structure of the A1A2-linker, the plant helix A2 should be brought in register with that found in the mammalian counterpart, to allow a correct recognition/binding of the CCAAT box: a Pro residue (Pro205 in At-NF-YA6) is completely conserved in plants, but not in mammalian NF-YA. Importantly, a similar Pro residue is also present in *fungi*, notably *S. cerevisiae* and *A. nidulans*, where Pro267 indeed contributes to orient the A2 helix of the HapB subunit, despite a different trajectory of the A1A2-linker (Figure 1F). However, EMSA experiments on a Pro205→Gly mutant (Figure S1C) indicate that the mutant retains a DNA-binding affinity comparable to that of the wt protein (Figure S4B). Thus, our analysis points out that this Pro residue, conserved in plants, may be responsible for favouring the correct functional orientation of the A2 helix towards the DNA in plant NF-YAs, but it is not essential for efficient DNA association and to stabilize the DNA bound complex, indicating a certain degree of structural adaptability of the A1A2-linker region.

In conclusion, At-L1L is a *bona fide* NF-YB subunit that can trimerize and bind to the CCAAT box in a mammalian-like mode. The structural data reported here highlight differences that help explain the molecular mechanisms of LEC1/L1L activity and provide the basis for further genetic and structural analysis of plants trimer/DNA interactions.

ACCESSION NUMBERS

The coordinates of the structure have been deposited in the Protein Data Bank with the accession code 5G49 .

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

This work was supported by UNIMI funding to N.G.

AUTHOR CONTRIBUTIONS

N.G., A.C.S., and D.S. performed the biochemical experiments; D.S., A.C.S. and M.N. crystallized the protein and solved the X-ray structure; N.G., M.N., and R.M. planned the experiments, analyzed the data, and wrote the manuscript.

ACKNOWLEDGMENTS

No conflict of interest declared.

REFERENCES

Braybrook, S.A. and Harada, J.J. (2008). LECs go crazy in embryo development. Trends Plant. Sci. **13:**624-630.

Calvenzani, V., Testoni, B., Gusmaroli, G., Lorenzo, M., Gnesutta, N., Petroni, K., Mantovani, R. and Tonelli, C. (2012). Interactions and CCAAT-binding of *Arabidopsis thaliana* NF-Y subunits. PLoS One **7:**e42902.

Dolfini, D., Zambelli, F., Pavesi, G., and Mantovani R. (2009). A perspective of promoter architecture from the CCAAT box. Cell Cycle **8:**4127-4137.

Huber, E.M., Scharf, D.H., Hortschansky, P., Groll, M. and Brakhage, A.A. (2012). DNA minor groove sensing and widening by the CCAAT-binding complex. Structure **20:**1757-1768.

Kwong, R.W., Bui, A.Q., Lee, H., Kwong, L.W., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (2003). LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. Plant Cell **15:**5-18.

Laloum, T., Baudin, M., Frances, L., Lepage, A., Billault-Penneteau, B., Cerri, M.R., Ariel, F., Jardinaud, M.F., Gamas, P., de Carvalho-Niebel, F. et al. (2014). Two CCAAT-box-binding transcription factors redundantly regulate early steps of the legume-rhizobia endosymbiosis. Plant J. **79:**757-768.

Lee, H., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (2003). *Arabidopsis* LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. Proc. Natl. Acad. Sci. USA **100:**2152-2156.

Mendes, A., Kelly, A.A., van Erp, H., Shaw, E., Powers, S.J., Kurup, S. and Eastmond, P.J. (2013). bZIP67 regulates the omega-3 fatty acid content of Arabidopsis seed oil by activating fatty acid desaturase3. Plant Cell **25:**3104-16.

Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., Vonrhein, C., Moras, D., Romier, C., Bolognesi, M. et al.. (2013). Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. Cell **152:**132-143.

Romier, C., Cocchiarella, F., Mantovani, R. and Moras, D. (2003). The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. J. Biol. Chem. **278:**1336-1345.

Sinha, S., Kim, I.S., Sohn, K.Y., de Crombrugghe, B. and Maity, S.N. (1996). Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. Mol. Cell. Biol. **16:**328-337.

Zemzoumi, K., Frontini, M., Bellorini, M. and Mantovani, R. (1999). NF-Y histone fold alpha1 helices help impart CCAAT specificity. J. Mol. Biol. **286:**327-737.

Yamamoto, A., Kagaya, Y., Toyoshima, R., Kagaya, M., Takeda, S. and Hattori, T. (2009). *Arabidopsis* NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. Plant J. **58:**843-856.

FIGURE LEGENDS

Figure 1. Structure and DNA-binding of At-L1L/NF-YC3.

(A) Ribbon diagram showing the histone fold dimer, and structural comparison of At-L1L /NF-YC3 with NF-YB/NF-YC (PDB: 4CSR), where the most divergent regions are highlighted by dashed circles.

(B) Electrostatic surface of the At-L1L/NF-YC3 dimer (top and side views) compared to that of NF-YB/NF-YC (PDB: 4AWL). Blue and red colors indicate positively and negatively charged regions, respectively. At-L1L His79 and Asp84 are indicated. NF-YA and the bound DNA are represented in ribbon and stick models, with the CCAAT box highlighted.

(C) Trimerization and DNA binding of At-L1L/NF-YC3 dimer with At-NF-YA6 was assessed by EMSA with Hsp70 CCAAT box DNA probe by addition of increasing doses of the respective NF-Y subunit(s) counterpart, as indicated. DNA binding mixes containing the HFD dimer (60 nM; lanes 2-5) or At-NF-YA6 (60 nM; lanes 7-10) were additioned with protein Dilution Buffer (DB; lanes 2, 7) or increasing amounts (20, 60 or 180 nM) of At-NF-YA6–6His or At-L1L/NF-YC3 (lanes 3-5; 8-10, respectively). As negative controls, At-NF-YA6–6His (lane 6) or At-L1L/NF-YC3 (lane 11) were incubated alone with the probe, at the highest concentration of the dose curve (180 nM). Lane 1: probe alone DNA binding mix without NF-Y subunits added. The At-NF-Y/DNA complex is indicated by an arrowhead. fp: free probe.

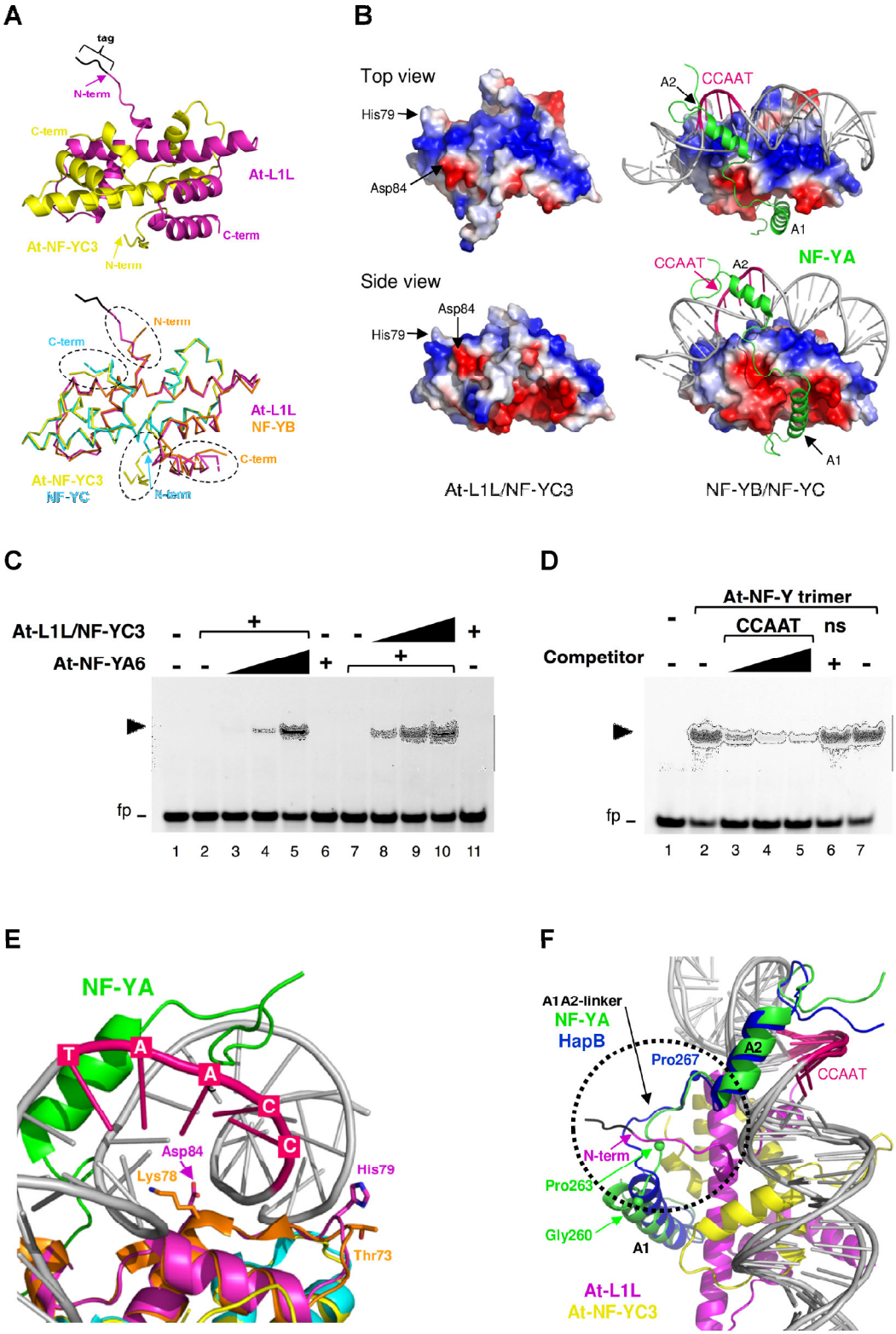
(D) CCAAT specificity of the At-L1L/NF-YC3/NF-YA6 NF-Y trimer/Hsp70 DNA complex was assessed in EMSA by addition of increasing doses of unlabeled CCAAT DNA or unrelated oligo (non-specific: ns) competitors. DNA binding mixes containing 20 nM labeled Hsp70 CCAAT probe alone (lane 1, 2, 7) or with increasing amounts of unlabeled oligo competitors (lane 3-5: Hsp70 CCAAT competitor: 100,

500, or 2000 nM; lane 6: ns competitor: 2000 nM) were additioned with At-NF-Y trimer subunits (lanes 2-7: 60 nM At-L1L/NF-YC3 and 180 nM At-NF-YA6-6His). Lane 1: probe alone DNA binding mix without NF-Y subunits added. The At-NF-Y/DNA complex is indicated by an arrowhead. fp: free probe.

(E) Position of the Asp84 and His79 residues. The superposition of the At-L1L/NF-YC3 dimer on the NF-Y/DNA complex (PDB: 4AWL) shows the position of At-L1L Asp84 and His79 relative to DNA and to the corresponding NF-YB residues Lys78 and Thr73.

(F) Structure of the At-L1L N-terminus. The superposition of the At-L1L/NF-YC3 dimer on the mammalian and *A. nidulans* NF-Y trimers (PDB: 4AWL and 4G92, respectively) shows the overlay of the At-L1L N-terminus relative to the NF-YA and HapB A1A2-linkers.

Gnesutta et al. Figure 1



SUPPLEMENTAL INFORMATION

Crystal structure of the *Arabidopsis thaliana* L1L/NF-YC3 histone-fold dimer reveals specificities of the LEC1 family of NF-Y subunits in plants.

Nerina Gnesutta[1], Dana Saad[1], Antonio Chaves-Sanjuan[1], Roberto Mantovani[1], Marco Nardini[1,*]

[1]Dipartimento di Bioscienze, Università degli Studi di Milano, Via Celoria 26, 20133, Milano, Italy.

***Correspondence: Marco Nardini** (marco.nardini@unimi.it)

SI Contents:

SUPPLEMENTAL FIGURES

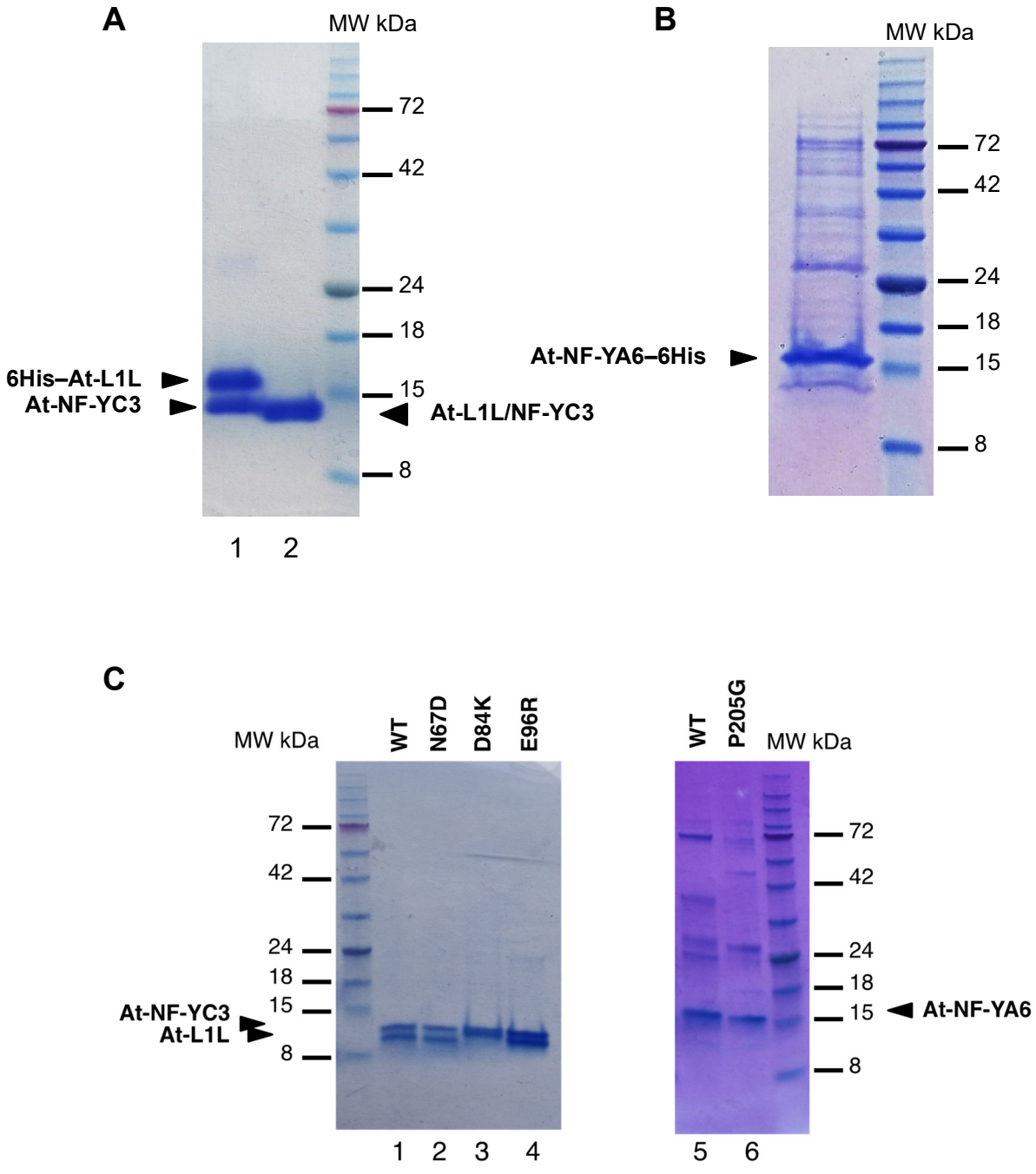


Figure S1

Figure S1. Coomassie stained SDS-PAGE gels (Bis-Tris – MES) of purified protein samples.

(A) Lane 1: IMAC purified *Arabidopsis thaliana* 6His–L1L/NF-YC3 (6 μg); lane 2: At-L1L/NF-YC3 protein after 6His-tag cleavage and GF purification, used for crystallization and EMSA analyses (6 μg). **(B)** IMAC purified At-NF-YA6–6His (6 μg) used for EMSA analyses. **(C)** Left panel: wt and mutant At-L1L/NF-YC3 purified protein dimers after 6His tag cleavage and GF purification (6 μg); Lane 1: wt At-L1L/NF-YC3 (WT); 2: At-L1LN67D/NF-YC3 (N67D); 3: At-L1LD84K/NF-YC3 (D84K); 4: At-L1LE96R/NF-YC3 (E96R). Right panel: IMAC purified wt and mutant At-NF-YA6-6His used for EMSA analyses (3 μg). Lane 1: wt At-NF-YA6 (WT); 2: At-NF-YA6P205G (P205G).

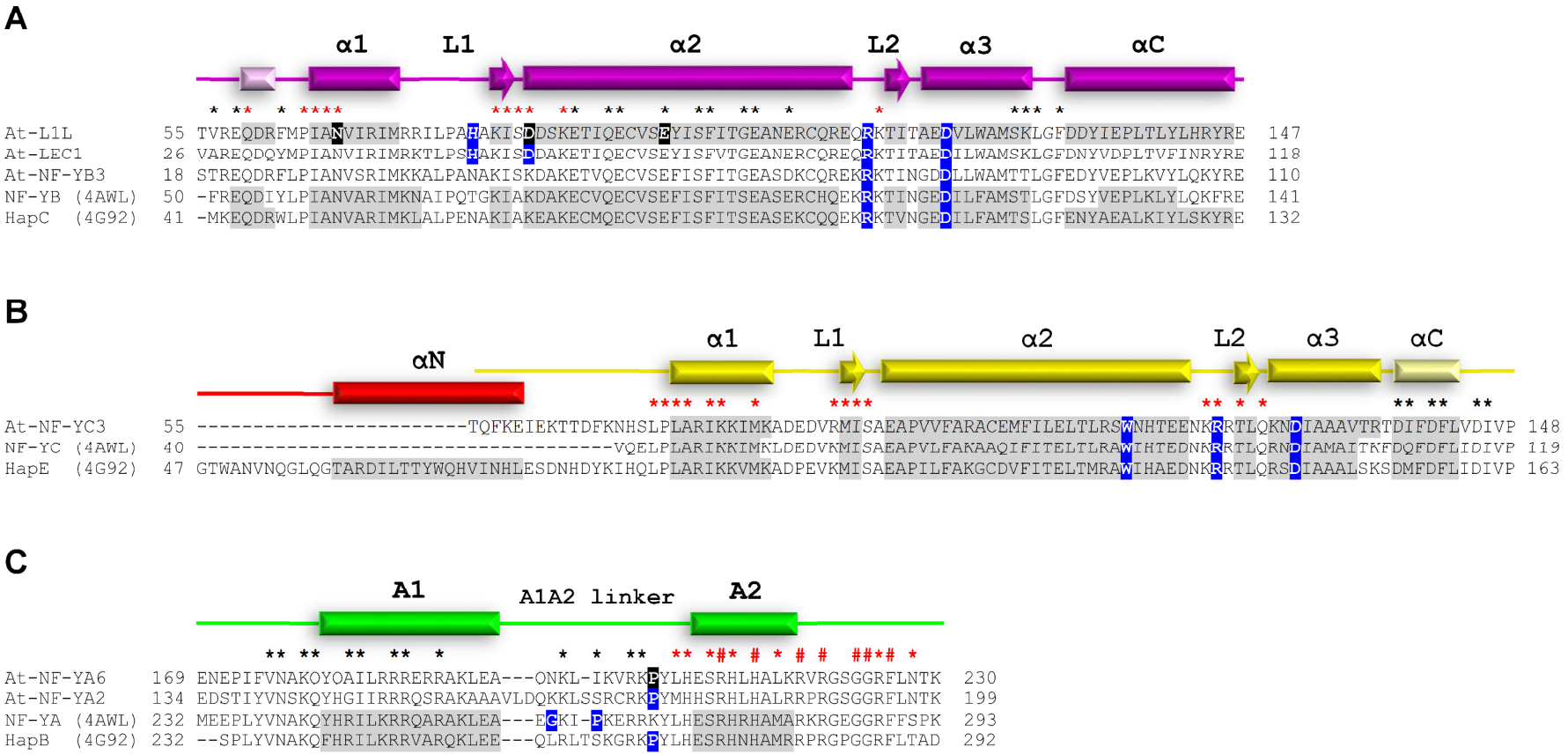

A
α1
L1
α2
L2
α3
αC
At-L1L 55 TVREQDRFMPIANVIRIMRRILPAHAKISDDSKETIQECVSEYISFITGEANERCQREQRKTITAEDVLWAMSKLGFDDYIEPLTLYLHRYRE 147
At-LEC1 26 VAREQDQYMPIANVIRIMRKTLPSHAKISDDAKETIQECVSEYISFVTGEANERCQREQRKTITAEDILWAMSKLGFDNYVDPLTVFINRYRE 118
At-NF-YB3 18 STREQDRFLPIANVSRIMKKALPANAKISKDAKETVQECVSEFISFITGEASDKCQREKRKTINGDDLLWAMTTLGFEDYVEPLKVYLQKYRE 110
NF-YB (4AWL) 50 -FREQDIYLPIANVARIMKNAIPQTGKIAKDAKECVQECVSEFISFITSEASERCHQEKRKTINGEDILFAMSTLGFDSYVEPLKLYLQKFRE 141
HapC (4G92) 41 -MKEQDRWLPIANVARIMKLALPENAKIAKEAKECMQECVSEFISFITSEASEKCQQEKRKTVNGEDILFAMTSLGFENYAEALKIYLSKYRE 132
B
αN
α1
L1
α2
L2
α3
αC
At-NF-YC3 55 ----------TQFKEIEKTTDFKNHSLPLARIKKIMKADEDVRMISAEAPVVFARACEMFILELTLRSWNHTEENKRRTLQKNDIAAAVTRTDIFDFLVDIVP 148
NF-YC (4AWL) 40 ----------VQELPLARIKKIMKLDEDVKMISAEAPVLFAKAAQIFITELTLRAWIHTEDNKRRTLQRNDIAMAITKFDQFDFLIDIVP 119
HapE (4G92) 47 GTWANVNQGLQGTARDILTTYWQHVINHLESDNHDYKIHQLPLARIKKVMKADPEVKMISAEAPILFAKGCDVFITELTMRAWIHAEDNKRRTLQRSDIAAALSKSDMFDFLIDIVP 163
C
A1
A1A2 linker
A2
At-NF-YA6 169 ENEPIFVNAKQYQAILRRRERRAKLEA---ONKL-IKVRKPYLHESRHLHALKRVRGSGGRFLNTK 230
At-NF-YA2 134 EDSTIYVNSKQYHGIIRRRQSRAKAAAVLDQKKLSSRCRKPYMHHSRHLHALRRPRGSGGRFLNTK 199
NF-YA (4AWL) 232 MEEPLYVNAKQYHRILKRRQARAKLEA---ECKI-EKERRKYLHESRHRHAMARKRGEGGRFFSPK 293
HapB (4G92) 232 --SPLYVNAKQFHRILKRRVARQKLEE---QLRLTSKGRKPYLHESRHNHAMRRPRGPGGRFLTAD 292

Figure S2

Figure S2. Structure-based sequence alignment of *A. thaliana* NF-Y core domains.

(A, B and C) *A. thaliana* NF-Y core domains are structurally aligned with human NF-Y (PDB: 4AWL) and with *A. nidulans* Hap (PDB: 4G92). The sequence alignment has been performed using the CLUSTALW program (http://www.ebi.ac.uk/Tools/msa/clustalw2/) and manually corrected based on the 3D structure comparison. For the alignment, only the aminoacids present in the crystal structures are considered. Secondary structure elements are shown as cartoon and labelled on the top of the alignment (magenta for At-L1L, yellow for At-NF-YC3, red for the N-terminus of *A. nidulans* HapE, green for NF-YA; 3_{10} helices are shown with light colors), and shaded in grey on the sequence. Black asterisks are associated to residues involved in NF-Y trimerization, while red asterisks indicate residues involved in DNA interactions. Red hash (#) symbol is associated to residues involved in DNA-base interactions. Residues described in the text of the article are highlighted in blue and mutated residues are highlighted in black.

(A) The At-L1L sequence is also aligned to At-LEC1 to highlight the conservation of the His79/Asp84 pair (His50/Asp55 in At-LEC1), not conserved in At-NF-YB3, taken as representative of all other "canonical" (non-LEC1-type) NF-YB subunits in *A. thaliana*. Note that orthologues of LEC1 and L1L identified in plant genomes analysed so far (Alemanno et al., 2008; Fambrini et al., 2006; Hilioti et al., 2014; Ito et al., 2011; Liang et al., 2014; Maillot et al., 2009; Salvini et al., 2012; Schellembaum et al., 2008; Xie et al., 2008; Yang et al., 2015; Yazawa et al., 2004, 2007) show conservation of these two amino acids, including in the evolutionarily distant *Selaginella moellendorffii* (Kirkbride et al., 2013; Saha et al., 2013). **(B)** The HFD region of At-NF-YC3 is aligned with human NF-YC and *A. nidulans* HapE. The secondary structure of the N-terminus of *A. nidulans* HapE is also shown as cartoon (red) to highlight its partial overlap with the N-terminus of At-NF-YC3. The full N-terminal sequence of At-NF-YC3 and human NF-YC is not reported since it is not present in the published structures. For a detailed information on the sequence alignment of this region in plant NF-YC refer to the following References: Cao et al., 2011; Feng et al., 2015; Li et al., 2016; Liang et al., 2014; Quach et al., 2015; Ripodas et al., 2015; Stephenson et al., 2007;

Thirumurugan et al., 2008; Zhang et al., 2015. **(C)** Core domains of At-NF-YA6 and At-NF-YA2 are aligned with human NF-YA and *A. nidulans* HapB. Secondary structure elements, referred to NF-YA, are shown as cartoon (green) and labelled on top of the alignment. They are also shaded in grey for both the available structures: NF-YA and HapB. Conservation of Pro205 in the A1A2-linker of At-NF-YA6 (Pro174 in At-NF-YA2) is highlighted, also conserved in other plant NF-YAs (see alignments in References indicated in B).

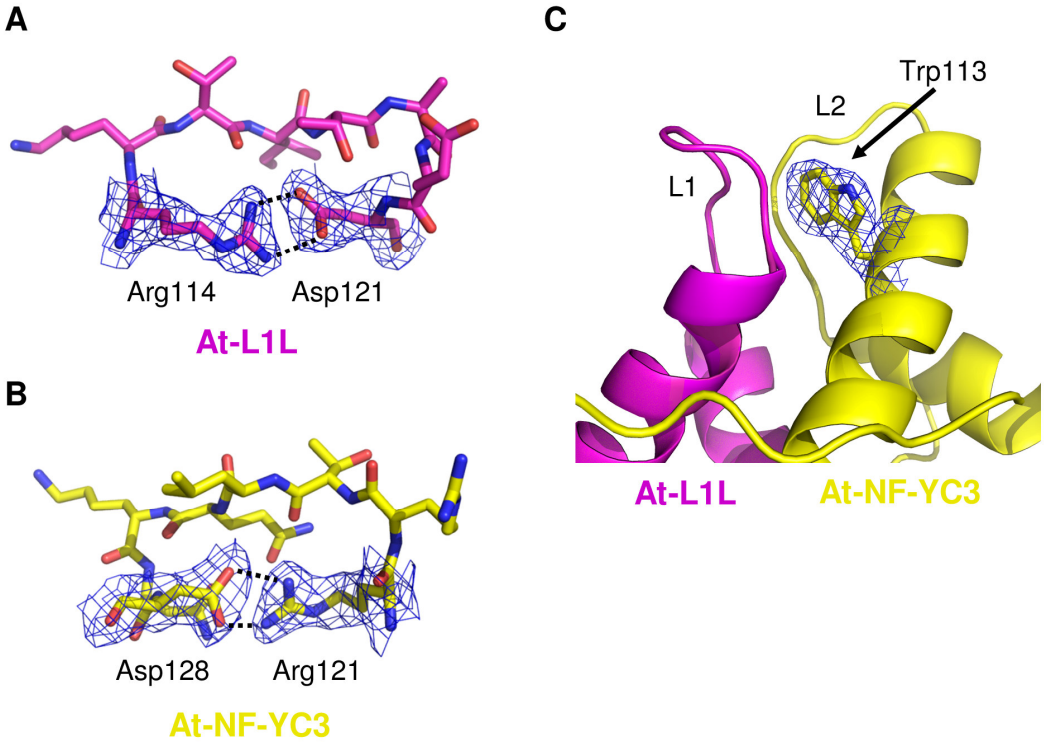


Figure S3. Structural features specific for HFD subunits.

(A, B) The Arg-Asp residues making the intra-chain bidentate pair (dashed lines) in At-L1L (Arg114-Asp121), and At-NF-YC3 (Arg121-Asp128) are shown in stick representation and labeled. **(C)** Similarly, the position of At-NF-YC3 Trp113, nestled at the interface between the L1/L2 loops of At-L1L and At-NF-YC3, is indicated. The refined 2Fo-Fc electron density map (contoured at 1.0 σ) around the residues is shown with a blue mesh.

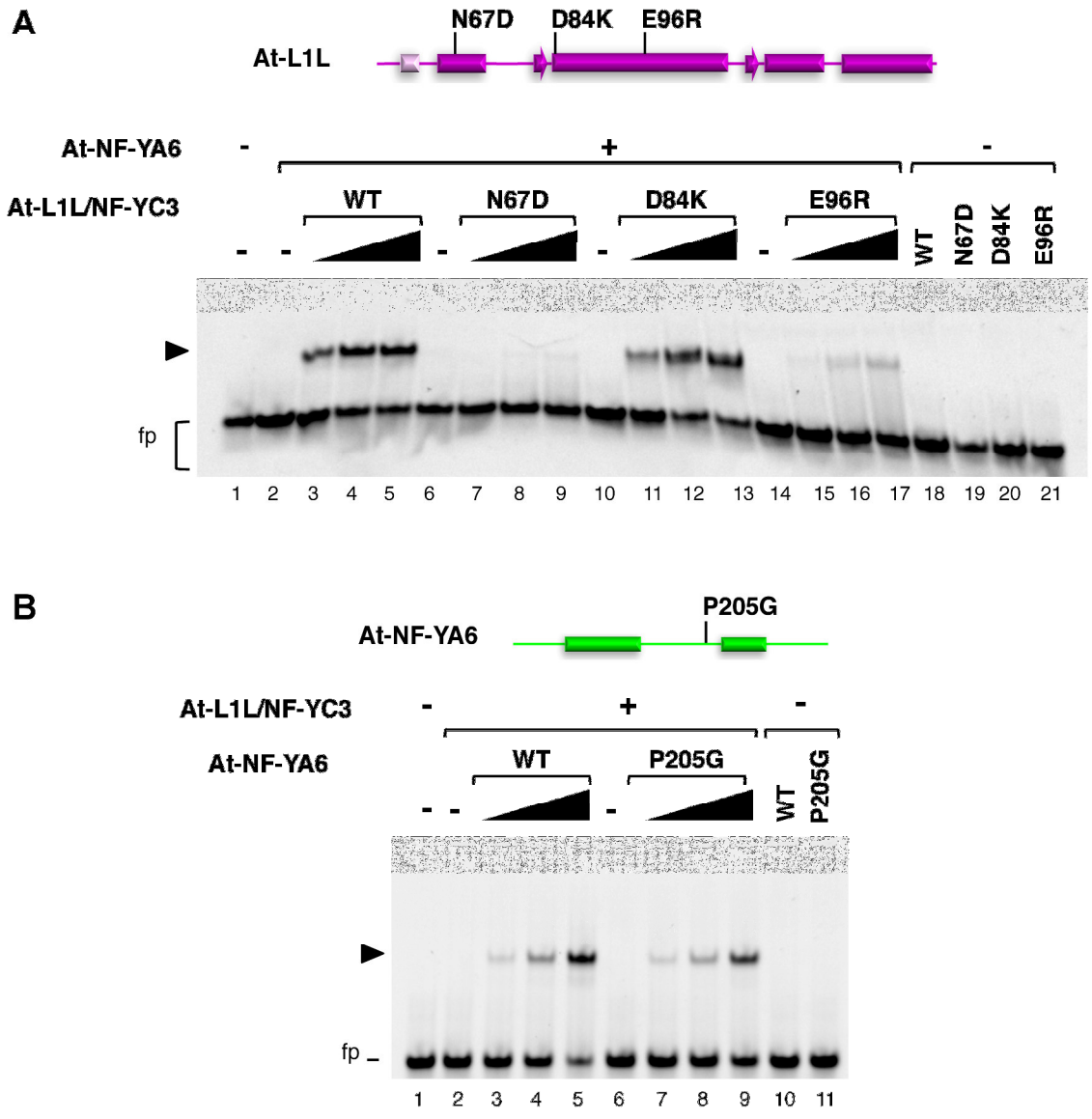

A
N67D
D84K
E96R
At-L1L
At-NF-YA6
At-L1L/NF-YC3
WT
N67D
D84K
E96R
fp
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
B
P205G
At-NF-YA6
At-L1L/NF-YC3
WT
P205G
fp
1 2 3 4 5 6 7 8 9 10 11

Figure S4

Figure S4. EMSA analysis of trimerization and DNA binding activity of At-L1L and At-NF-YA6 mutants.

(A) At-L1L/NF-YC3 HFD mutants. Relevant At-L1L residues mutant dimers were tested for trimerization with At-NF-YA6 and DNA binding with the Hsp70 CCAAT box probe by EMSA. At-L1L/NF-YC3 wt or mutant HFD dimers were incubated at increasing concentrations (20, 60 or 180 nM) with the Hsp70 probe in the presence (+) of At-NF-YA6 (60 nM): lanes 3-5 At-L1L/NF-YC3 wt (WT); lanes 7-9 At-L1LN67D/NF-YC3 (N67D); lanes 11-13 At-L1LD84K/NF-YC3 (D84K); lanes 15-17 At-L1LE96R/NF-YC3 (E96R). As controls, the HFD dimers containing wt or mutant At-L1L, as indicated, were incubated with the probe at the highest concentration of the dose curve (180 nM), without (-) At-NF-YA6 addition (lanes 18-21). Lane 1: probe alone, without protein additions; lanes 2, 6, 10, 14: probe incubated with At-NF-YA6 (60 nM), without any HFD dimer addition. The At-NF-Y/DNA complex is indicated by an arrowhead. fp: free probe.

(B) At-NF-YA6 P205G mutant was tested for trimerization with At-L1L/NF-YC3 and DNA binding with the Hsp70 CCAAT box probe by EMSA. wt or mutant At-NF-YA6 was incubated at increasing concentrations (20, 60 or 180 nM) with the Hsp70 probe in the presence (+) of At-L1L/NF-YC3 HFD dimers (60 nM). Lanes 3-5: At-NF-YA6 wt (WT); lanes 7-9 At-NF-YA6P205G (P205G). As controls, wt or mutant At-NF-YA6, as indicated, were incubated with the probe at the highest concentration of the dose curve (180 nM), without (-) At-L1L/NF-YC3 addition (lanes 10,11). Lane 1: probe alone, without protein additions; lanes 2 and 6: probe incubated with At-L1L/NF-YC3 alone (60 nM), without At-NF-YA6 addition. The At-NF-Y/DNA complex is indicated by an arrowhead. fp: free probe.

On top of panels A and B, a scheme of the subunit with the mutated residue positions is indicated.

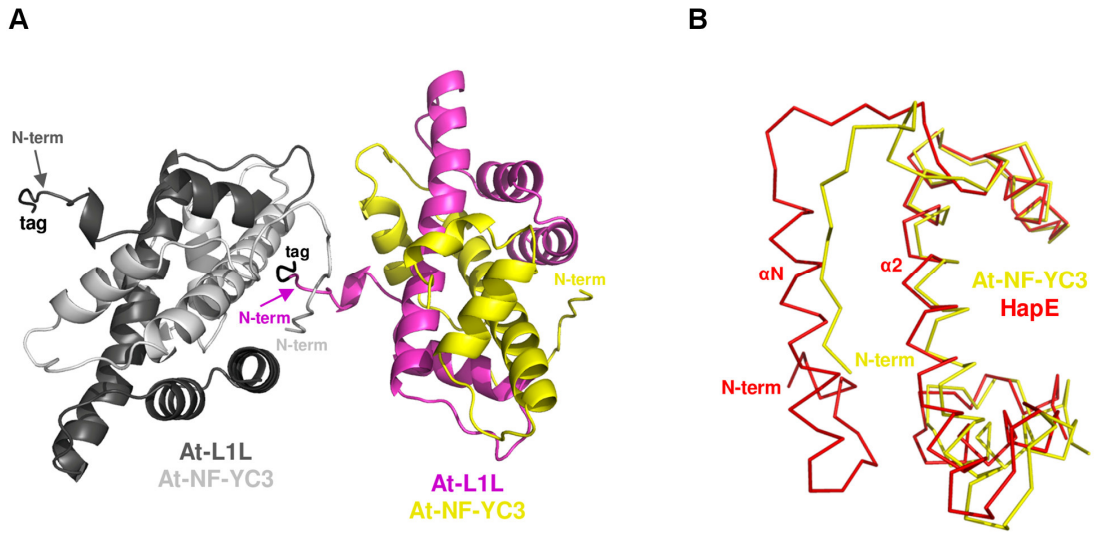


Figure S5. Structure of the N-termini of At-L1L/NF-YC3.

(A) The packing of the At-L1L N-terminal region (magenta) against the N-terminal region of an At-NF-YC3 subunit, belonging to an adjacent asymmetric unit (light grey), is shown. **(B)** Structural overlay of At-NF-YC3 (yellow) with *A. nidulans* HapE (red) (PDB: 4G92). The N-terminus of both proteins is labeled, and the αN and α2 helices of HapE highlighted.

SUPPLEMENTAL TABLES

Supplemental Table 1. *Diffraction data and structure refinement statistics*

Diffraction data	
Space group	$C222_1$
Cell parameters	
a, b, c (Å)	68.3, 81.7, 96.0
α, β, γ (°)	90.0, 90.0, 90.0
Resolution	52.42-2.30 (2.42-2.30)[a]
Observed reflections	106,408
Unique reflections	12,288
Completeness (%)	100 (100)
Average I/σ(I)	20.9 (5.1)
Average redundancy	8.7 (9.1)
R_{merge} (%)[b]	7.3 (49.6)
Structure refinement	
R_{factor} / R_{free} (%)[c]	16.2 / 20.0
N° of atoms	
Protein	1,592
Ca^{2+}	1
Water	109
Acetate	9
Average B factor (Å^2)	
Protein atoms	48.0
Ca^{2+}	66.4
Water	52.8
Acetate	68.5
RMS deviations	
Bond lengths (Å)	0.008
Bond angles (°)	0.939
Ramachandran plot (%)	
Most favored	97.0
Allowed	3.0

[a] Values in parentheses refer to the last resolution shell.
[b] $R_{merge}=\sum|I-<I>|/\sum I\times 100$, where I is intensity of a reflection and <I> is the average intensity.
[c] $R_{factor}=\sum|F_o-F_c|/\sum F_o\times 100$. R_{free} is calculated from 5% randomly selected data for cross-validation.

SUPPLEMENTAL MATERIALS AND METHODS

Cloning, Expression and Purification

The *Arabidopsis thaliana* 6His–L1L/NF-YC3 soluble HFD dimer (At-L1L aa 26-118; At-NF-YC3 aa 55-148) was produced by co-expression in *E. coli* BL21(DE3) and purified by Ion Metal Affinity Chromatography (IMAC) as described previously (Calvenzani et al., 2012). Following proteolytic cleavage of the 6His-tag (Thrombin CleanCleave Kit, Sigma-Aldrich), the HFD dimer was further purified by gel filtration (GF) chromatography in buffer B (10 mM Tris-HCl pH 8.0, 400 mM NaCl, 2 mM DTT) and concentrated for crystallization (Figure S1A). 10% Glycerol was added for storage of purified protein used in EMSA analyses. Expression vectors encoding At-NF-Y mutant subunits 6His-L1L (aa 26-118) with residue N67 mutated to D (At-L1LN67D), residue D84 mutated to K (At-L1LD84K), and residue E96 mutated to R (At-L1LE96R) were obtained by site directed mutagenesis (Quickchange, Stratagene) using as template the relative wt 6His-L1L expression vector (Calvenzani et al., 2012). At-L1L mutants were produced by coexpression with At-NF-YC3 and purified by IMAC as described for the wt protein (Calvenzani et al., 2012), and GF purified after proteolytic cleavage of the 6His tag (Figure S1C).

The cDNAs encoding for At-NF-YA6 (aa 170-237 with a starting M residue) wt, or with residue P205 mutated to G (At-NF-YA6P205G) were obtained by gene synthesis (Eurofins Genomics) and subcloned in pmcnEATCH (Diebold et al., 2011) to obtain At-NF-YA6–6His C-terminal fusions. All plasmid constructs were verified by sequencing. At-NF-YA6–6His wt and mutant were expressed in *E. coli* BL21(DE3) by IPTG induction (0.4 mM IPTG for 4h at 25 °C) and purified by IMAC (HisSelect, SIGMA-Aldrich) in buffer A (10 mM Tris-HCl pH 8.0, 400 mM NaCl, 2 mM $MgCl_2$, 5 mM imidazole) (Figures S1B and S1C). The purified proteins were eluted in Buffer A containing 100 mM imidazole, and dialysed against Buffer B additioned with 10% glycerol.

Electrophoretic Mobility Shift Assays

EMSA analyses were performed essentially as described in (Calvenzani et al., 2012; Cao et al., 2014). Heterotrimer formation and CCAAT-box DNA-binding of GF purified At-L1L/NF-YC3 dimers was assessed with Cy5-labeled oligos, by addition of IMAC purified At-NF-YA6–6His. In dose response NF-Y trimerization tests DNA binding mixes containing or not the indicated NF-Y subunits (HFD dimers or At-NF-YA6–6His; 60 nM) were prepared with 5'-labeled 31 bp oligo CCAAT probe derived from human HSP70 promoter (Cy5-cttctgagCCAATcaccgagctcgatgaggc) (20 nM Hsp70 CCAAT probe, 12 mM Tris-HCl pH 8, 50 mM KCl, 62.5 mM NaCl, 0.5 mM EDTA, 5 mM MgCl2, 2.5 mM DTT, 0.2 mg/ml BSA, 5% glycerol, 6.25 ng/µl polydA-dT), and additioned with increasing amounts of At-NF-YA6–6His or At-L1L/NF-YC3 (20, 60 or 180 nM) respectively, or protein Dilution Buffer (DB: Buffer B containing 10% glycerol, 0.1 mg/ml BSA), as indicated in Figure 1C. As negative controls, At-NF-YA6–6His or At-L1L/NF-YC3 were incubated alone with the probe, at the highest concentration of the dose curve (180 nM). After 30 min incubation at 30 °C, binding reactions were loaded on 6% acrylamide gels and separated by electrophoresis in 0.25X TBE. A similar experimental set-up was used for At-L1L and At-NF-YA6 mutants EMSA analyses shown in Figure S4. For competition assays, DNA binding reactions in the presence of increasing amounts of unlabelled Hsp70 CCAAT (ttctgagCCAATcaccgagctcgat) (100, 500, or 2000 nM) or unrelated sequence from human HOXB4 promoter (ns competitor: ttaggcgcccacgtgatcctccga) (2000 nM) oligonucleotide competitors or TE buffer, as indicated in Figure 1D, were additioned with At-NF-Y proteins (60 nM At-L1L/NF-YC3 and 180 nM At-NF-YA6–6His) and incubated for 30 min at 30 °C. Fluorescence gel images were obtained with a Chemidoc MP system (Bio-Rad) with ImageLab software.

Crystallization and Data Collection

Crystallization trials were set up manually in vapour diffusion at 4 °C and 21 °C, after concentration steps that brought the At-L1L/NF-YC3 sample to 10.5 mg/ml. This value was selected since the human NF-YB/NF-YC dimer was successfully crystallized in the concentration

range of 10-14 mg/ml (Romier et al., 2003). The best results were obtained in the following crystallization conditions: 10% w/v PEG 8,000, 100 mM HEPES-NaOH pH 7.5, 200 mM Calcium acetate. Prior to data collection at 100 K, crystals were transferred to the following cryo-solutions: 15% w/v PEG 8,000, 100 mM HEPES-NaOH pH 7.5, 200 mM Calcium acetate, 30% (v/v) glycerol.

The At-L1L/NF-YC3 crystals were tested for X-ray diffraction at the beamline ID23-1 of the European Synchrotron Radiation Facility (ESRF) of Grenoble, France (www.esrf.fr). They diffracted up to 2.3 Å and belong to the C2221 space group, with unit cell parameters: a = 68.3 Å, b = 81.7 Å, c = 96.0 Å and $\alpha = \beta = \gamma = 90.0°$. The Matthews coefficient (VM) is 3.57 $Å^3$ Da^{-1}, with 65.5% solvent content, assuming one At-L1L/NF-YC3 dimer (18,788 Da) in the asymmetric unit. All collected data were reduced and scaled using MOSFLM (Leslie, 2003) and SCALA (Evans, 2006) and programs of the CCP4 package (Collaborative Computational Project, Number 4, 1994). Data processing statistics are reported in Table S1.

Structure Determination and Refinement

The At-L1L/NF-YC3 crystal structure was solved by molecular replacement using the program Phaser (Storoni et al., 2004), with the crystal structure of NF-YB/NF-YC (PDB: 4CSR) as a search model. One protein dimer was successfully located in the crystal asymmetric unit. The positioned model was subjected to rigid-body and restrained refinement using the program Refmac (Murshudov et al., 1997). A random set comprising 5% of the data was omitted from refinement for Rfree calculation. The amino acid sequence of the model was modified to match the correct sequence and manually fitted to the electron density using the program Coot (Emsley and Cowtan, 2004). A final set of restrained refinement cycles were performed using the programs Phenix (Adams et al., 2010).

The final refined At-L1L/NF-YC3 structure (Rfactor = 16.2%, Rfree = 20.0%) contains 191 protein residues (At-L1L: 55-147 and GSHM residues from the N-terminal tag; At-NF-YC3: 55-

148) out of 192 total residues, 109 water molecules, 3 acetate molecules, and one Ca^{2+} ion, with good stereochemical parameters, as calculated by the programs Molprobity (Chen et al., 2010). Refinement statistics are shown in Table S1. Structure images were generated using Pymol (The PyMOL Molecular Graphics System, Schrödinger, LLC).

SUPPLEMENTAL REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. **66:**213-221.

Alemanno, L., Devic, M., Niemenak, N., Sanier, C., Guilleminot, J., Rio, M., Verdeil, J.L. and Montoro, P. (2008). Characterization of leafy cotyledon1-like during embryogenesis in *Theobroma cacao* L. Planta **227:**853-866.

Calvenzani, V., Testoni, B., Gusmaroli, G., Lorenzo, M., Gnesutta, N., Petroni, K., Mantovani, R. and Tonelli, C. (2012). Interactions and CCAAT-binding of *Arabidopsis thaliana* NF-Y subunits. PLoS One **7:**e42902.

Cao, S., Kumimoto, R.W., Siriwardana, C.L., Risinger, J.R. and Holt, B.F. III. (2011). Identification and Characterization of NF-Y Transcription Factor Families in the Monocot Model Plant *Brachypodium distachyon*. PLoS ONE **6:**e21805.

Cao, S., Kumimoto, R.W., Gnesutta, N., Calogero, A.M., Mantovani, R. and Holt B.F. 3rd. (2014). A distal CCAAT/NUCLEAR FACTOR Y complex promotes chromatin looping at the FLOWERING LOCUS T promoter and regulates the timing of flowering in Arabidopsis. Plant Cell **26:**1009-1017.

Chen, V.B., Arendall III, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. **66:**12-21.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. **50:**760-763.

Diebold, M.L., Fribourg, S., Koch, M., Metzger, T. and Romier, C. (2011). Deciphering correct strategies for multiprotein complex assembly by coexpression: application to complexes as large as the histone octamer. J. Struct. Biol. **175:**178-188.

Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. **60:**2126-2132.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. **62:**72-82.

Fambrini, M., Durante, C., Cionini, G., Geri, C., Giorgetti, L., Michelotti, V., Salvini, M. and Pugliesi, C. (2006). Characterization of LEAFY COTYLEDON1-LIKE gene in *Helianthus annuus* and its relationship with zygotic and somatic embryogenesis. Development Genes and Evolution **216:**253-264.

Feng, Z.J., He, G.H., Zheng, W.J., Lu, P.P., Chen, M., Gong, Y.M., Ma, Y.Z. and Xu, Z.S. (2015). Foxtail Millet NF-Y families: genome-wide survey and evolution analyses identified two functional genes important in abiotic stresses. Front. Plant Sci. **6:**1142.

Hilioti, Z., Ganopoulos, I., Bossis, I. and Tsaftaris, A. (2014). LEC1-LIKE paralog transcription factor: how to survive extinction and fit in NF-Y protein complex. Gene **543:**220-233.

Kirkbride, R.C., Fischer, R.L. and Harada, J.J. (2013). LEAFY COTYLEDON1, a key regulator of seed development, is expressed in vegetative and sexual propagules of *Selaginella moellendorffii*. PLoS One **8:**e67971.

Ito, Y., Thirumurugan, T., Serizawa, A., Hiratsu, K., Ohme-Takagi, M. and Kurata, N. (2011). Aberrant vegetative and reproductive development by overexpression and lethality by silencing of OsHAP3E in rice. Plant Science **181:**105-110.

Leslie, A.G.M. (2003). MOSFLM User Guide, Mosflm Version 6.2.3, Cambridge, UK: MRC Laboratory of Molecular Biology.

Li, S., Li, K., Ju, Z., Cao, D., Fu, D., Zhu, H., Zhu, B. and Luo, Y. (2016). Genome-wide analysis of tomato NF-Y factors and their role in fruit ripening. BMC Genomics **17:**36.

Liang, M., Yin, X., Lin, Z., Zheng, Q., Liu, G. and Zhao, G. (2014). Identification and characterization of NF-Y transcription factor families in Canola (*Brassica napus* L.). Planta **239:**107-126

Maillot, P., Lebel, S., Schellenbaum, P., Jacques, A. and Walter, B. (2009). Differential regulation of SERK, LEC1-like and pathogenesis-related genes during indirect secondary somatic embryogenesis in grapevine. Plant Physiology and Biochemistry **47:**743-752.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. **53:**240-255.

Quach, T.N., Nguyen, H.T., Valliyodan, B., Joshi, T., Xu, D. and Nguyen, H.T. (2015). Genome-wide expression analysis of soybean NF-Y genes reveals potential function in development and drought response. Mol. Genet. Genomics **290:**1095-115.

Rípodas, C., Castaingts, M., Clúa, J., Blanco, F. and Zanetti, M.E. (2015). Annotation, phylogeny and expression analysis of the nuclear factor Y gene families in common bean (*Phaseolus vulgaris*). Front. Plant Sci. **5:**761.

Romier, C., Cocchiarella, F., Mantovani, R. and Moras, D. (2003). The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. J. Biol. Chem. **278:**1336-1345.

Saha, J., Gupta, K. and Gupta, B. (2013). *In silico* characterization and evolutionary analyses of CCAAT binding proteins in the lycophyte plant *Selaginella moellendorffii* genome: a growing comparative genomics resource. Comput. Biol. Chem. **47:**81-88.

Salvini, M., Sani, E., Fambrini, M., Pistelli, L., Pucciariello, C. and Pugliesi, C. (2012). Molecular analysis of a sunflower gene encoding an homologous of the B subunit of a CAAT binding factor. Mol. Biol. Rep. **39:**6449-6465.

Schellenbaum, P., Jacques, A., Maillot, P., Bertsch, C., Mazet, F., Farine, S. and Walter, B. (2008). Characterization of VvSERK1, VvSERK2, VvSERK3 and VvL1L genes and their

expression during somatic embryogenesis of grapevine (*Vitis vinifera L.*). Plant Cell. Rep. **27:**1799-1809.

Stephenson, T.J., McIntyre, C.L., Collet, C. and Xue, G.P. (2007). Genome-wide identification and expression analysis of the NF-Y family of transcription factors in *Triticum aestivum*. Plant Mol. Biol. **65:**77-92.

Storoni, L.C., McCoy, A.J. and Read, R.J. (2004). Likelihood-enhanced fast rotation functions. Acta Crystallogr. D Biol. Crystallogr. **60:**432-438.

Thirumurugan, T., Ito, Y., Kubo, T., Serizawa, A. and Kurata, N. (2008). Identification, characterization and interaction of HAP family genes in rice. Mol. Genet. Genomics **279:**279-289.

Xie, Z., Li, X., Glover, B.J., Bai, S., Rao, G.Y., Luo, J. and Yang, J. (2008). Duplication and functional diversification of HAP3 genes leading to the origin of the seed-developmental regulatory gene, LEAFY COTYLEDON1 (LEC1), in nonseed plant genomes. Molecular biology and evolution **25:**1581-1592.

Yang, S., Wang, Y., Yin, H., Guo, H., Gao, M., Zhu, H. and Chen, Y. (2015). Identification and characterization of NF-YB family genes in tung tree. Mol. Genet. Genomics **290:**2187-98.

Yazawa, K., Takahata, K. and Kamada, H. (2004). Isolation of the gene encoding Carrot leafy cotyledon1 and expression analysis during somatic and zygotic embryogenesis. Plant Physiology and Biochemistry **42:**215-223.

Yazawa, K. and Kamada, H. (2007). Identification and characterization of carrot HAP factors that form a complex with the embryo-specific transcription factor CLEC1. J. Exp. Bot. **58:**3819-3828.

Zhang, F., Han, M., Lv, Q., Bao, F. and He, Y. (2015). Identification and expression profile analysis of NUCLEAR FACTOR-Y families in *Physcomitrella patens*. Front. Plant Sci. **6:**642.